

Statistical Mechanics Approach to Sparse Noise Denoising

Mikko Vehkaperä^{1,2}, Yoshiyuki Kabashima³, Saikat Chatterjee¹

¹KTH Royal Institute of Technology and the ACCESS Linnaeus Center, SE-100 44, Stockholm, Sweden

²Aalto University, P.O. Box 11000, FI-00076 AALTO, Finland

³Tokyo Institute of Technology, Yokohama, 226-8502, Japan

E-mails: mikkov@kth.se, kaba@dis.titech.ac.jp, sach@kth.se

Abstract—Reconstruction fidelity of sparse signals contaminated by sparse noise is considered. Statistical mechanics inspired tools are used to show that the ℓ_1 -norm based convex optimization algorithm exhibits a phase transition between the possibility of perfect and imperfect reconstruction. Conditions characterizing this threshold are derived and the mean square error of the estimate is obtained for the case when perfect reconstruction is not possible. Detailed calculations are provided to expose the mathematical tools to a wide audience.

Index Terms—sparse signals and noise, replica method, statistical mechanical analysis

I. INTRODUCTION

Sparse signal estimation for linear underdetermined systems has attracted wide interest in signal processing community during the recent years. This is not surprising since the general class of sparse problems is encountered in many applications, such as, linear regression [1], multimedia [2], [3], and compressive sampling (CS) [4], [5], to name just a few.

The present paper considers a CS setup where the sparse vector $\mathbf{x} \in \mathbb{R}^N$ is observed via noisy linear measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ represents the compressive ($M < N$) sampling system and $\mathbf{y} \in \mathbb{R}^M$ is the observed vector. The measurement errors are captured by the additive noise vector $\mathbf{w} \in \mathbb{R}^M$. The task is then to reconstruct \mathbf{x} from \mathbf{y} , given \mathbf{A} . Typically, detailed information about the statistics of \mathbf{x} and \mathbf{w} is not available to the system designer apart from some general knowledge, such as, “signal and noise are sparse”.

A prominent approach for finding a sparse solution to (1) is by solving a (convex) optimization problem of the form

$$\hat{\mathbf{x}}_\lambda = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \{C_{\mathbf{y}, \mathbf{A}}(\mathbf{x}) + \lambda \|\mathbf{x}\|_1\}, \quad (2)$$

where $\|\mathbf{x}\|_1 = \sum_n |x_n|$ denotes the standard ℓ_1 -norm. The non-negative cost function $C_{\mathbf{y}, \mathbf{A}}(\mathbf{x})$ that may depend on the realizations of \mathbf{y} and \mathbf{A} is typically chosen so that the solution to (2) can be obtained using standard convex optimization tools like `cvx` [6]. In addition to the choice of $C_{\mathbf{y}, \mathbf{A}}(\mathbf{x})$, the solution also depends on the regularization parameter λ . In general, finding the optimal value of λ is not a trivial task.

For the case of (dense) Gaussian noise, the standard approach is to set $C_{\mathbf{y}, \mathbf{A}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$, reducing (2) to the so-called LASSO estimator [7]. It is well-known that for non-zero noise variance, the solution obtained through LASSO is not exact. However, if the noise has some structure, like sparsity, perfect reconstruction could again be feasible. Indeed, in many applications the noise \mathbf{w} can be considered to be sparse, some examples of which are: impulsive noise [8], salt-and-pepper noise in an image, a sensor scenario where few measurements are corrupted but the other ones are good [9], and dictionary learning with sparse noise [10].

Inspired by the above practical examples, let us consider a setup where both \mathbf{x} and \mathbf{w} are sparse. To take into account the sparseness of noise, we set

$$C_{\mathbf{y}, \mathbf{A}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1, \quad (3)$$

and note that this choice makes (2) a convex optimization problem. We then ask the following questions:

- 1) Given that the signal and noise are sparse with some prior distributions (that we do not know), what values of M and N allow perfect reconstruction of \mathbf{x} ?
- 2) What is the mean square error (MSE) of the sparse estimate of \mathbf{x} outside of this region?

We answer these questions in the large system limit (LSL) by using mathematical tools developed in statistical mechanics. The key technique is the *replica trick*¹ that can be used to assess the *free energy* of a disordered system [11]. This is particularly useful in equilibrium statistical mechanics since the large-scale behavior of the system can be inferred from it. Similar notions have been made in information theory [12]–[15] and in CS [16]–[19], where quantities like mutual information and MSE play the role of thermodynamic variables.

II. PROBLEM FORMULATION AND METHODS

Consider the set of noisy measurements (1) and assume that both the signal and the noise are sparse random vectors (RVs). We define a parametrized mixture distribution

$$p(z; \rho, \sigma^2) = (1 - \rho)\delta(z) + \rho g_z(0, \sigma^2), \quad (4)$$

¹The most serious drawback of the replica method is that some of its steps are still lacking formal proof. Hence, it can be considered to be at most a “semi-rigorous” analytical tool. Regardless, the replica trick is routinely used in equilibrium statistical mechanics and its predictions are often verified by experiments. We shall also demonstrate this later in the paper.

where $g_z(\mu, \sigma^2) = e^{-(z-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2}$, and let the elements of \mathbf{x} (resp. \mathbf{w}) be independently and identically distributed (IID) according to $p(x; \rho_x, \sigma_x^2)$ (resp. $p(w; \rho_w, \sigma_w^2)$). Note that $\rho \in [0, 1]$ in (4) gives the fraction of non-zero Gaussian elements in the vector whereas σ^2 is the variance of them. The measurement process is taken to be random so that the elements of \mathbf{A} are IID with density² $g_A(0, 1/N)$. For future convenience, we denote the ratio between the number of observables and the unknown parameters by $\alpha = M/N$. To enable the use of statistical mechanics tools, we next write the original problem in a probabilistic framework.

Let us consider the optimization problem (2) with the ℓ_1 -cost (3). Assume the system is in the LSL $M, N \rightarrow \infty$, where the compression ratio $\alpha = M/N$ and the density of the signal and noise ρ_x, ρ_w remain constant. Let the postulated prior of \mathbf{x} be proportional to the Laplace distribution, namely, $q_{\beta, \lambda}(\mathbf{x}) \propto e^{-\beta\lambda\|\mathbf{x}\|_1}$, where $\beta \geq 0$ (inverse temperature in statistical mechanics). The postulated distribution of the measurement process has the same form, that is, $q_{\beta}(\mathbf{y} | \mathbf{A}, \mathbf{x}) \propto e^{-\beta\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1}$, and the (mismatched) conditional mean estimator of \mathbf{x} reads by definition

$$\langle \mathbf{x}; \lambda \rangle_{\beta} = \frac{1}{Z_{\beta}(\mathbf{y}, \mathbf{A}; \lambda)} \int \mathbf{x} q_{\beta}(\mathbf{y} | \mathbf{A}, \mathbf{x}) q_{\beta, \lambda}(\mathbf{x}) d\mathbf{x}, \quad (5)$$

where $Z_{\beta}(\mathbf{y}, \mathbf{A}; \lambda) = \int e^{-\beta(\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_1 + \lambda\|\mathbf{x}\|_1)} d\mathbf{x}$. Then, the zero temperature estimate $\hat{\mathbf{x}}_{\lambda} = \langle \mathbf{x}; \lambda \rangle_{\beta \rightarrow \infty}$, is the solution to the original optimization problem defined by (2) and (3).

A. Replica Method

The key for finding the statistical properties of the reconstruction (5) is the relatively innocent looking normalization factor $Z_{\beta}(\mathbf{y}, \mathbf{A}; \lambda)$, that acts as the *partition function* of the system. Based on the standard approach in statistical mechanics, our goal should then be to assess the normalized *free energy* $f_{\beta}(\mathbf{y}, \mathbf{A}; \lambda) = \frac{1}{\beta N} \ln Z_{\beta}(\mathbf{y}, \mathbf{A}; \lambda)$, when $N \rightarrow \infty$ and obtain the desired statistical properties from it. This formulation is, however, problematic since f depends on the observations \mathbf{y} and the measurement process \mathbf{A} . A way around is to consider the averaged quantity $f_{\beta}(\lambda) = \mathbb{E} f_{\beta}(\mathbf{y}, \mathbf{A}; \lambda)$ instead. Unfortunately, the problem of computing an expectation over a logarithm is in general very difficult. Hence, we reformulate $f_{\beta}(\lambda)$ in the zero temperature limit as

$$f(\lambda) = \lim_{\beta, N \rightarrow \infty} \frac{1}{\beta N} \lim_{u \rightarrow 0^+} \frac{\partial}{\partial u} \ln \mathbb{E}\{[Z_{\beta}(\mathbf{y}, \mathbf{A}; \lambda)]^u\}, \quad (6)$$

and remark that so-far our treatment has been rigorous if the normalized free energy exists in the LSL (we do not dwell on this technical point due to space constraints). Unfortunately, obtaining an expression for (6) is still difficult. This is when the replica trick comes in handy.

Replica trick. Consider the normalized free energy in (6) and assume that we can change the order of the limits. Postulate

²It is a common practice in CS to normalize the columns of \mathbf{A} instead of the rows. The former can be obtained here by redefining the signal power to be $\alpha\sigma_x^2$. This, however, has no effect on the perfect recovery conditions.

further that inside the logarithm we may write

$$[Z_{\beta}(\mathbf{y}, \mathbf{A}; \lambda)]^u = \int \prod_{a=1}^u e^{-\beta(\|\mathbf{y} - \mathbf{A}\mathbf{x}^a\|_1 + \lambda\|\mathbf{x}^a\|_1)} d\mathbf{x}^a, \quad (7)$$

and take the limit $u \rightarrow 0^+$ outside of it, after we have calculated the expectations for integer u .

The most debatable step in the replica trick is the assumption that the variable u (number of replicas) can be first treated as a non-negative integer and then extended to the set of real numbers. There is no rigorous mathematical proof that guarantees this can be done. Numerous results both in physics and engineering show, however, that the predictions of the replica method tend to be accurate when compared to numerical experiments. For the rest of the paper we assume thus that the replica trick is valid and verify its predictions in the end using Monte Carlo simulations.

The general scheme for the replica analysis consists of first assessing (6) using the above replica trick and then identify the parameters that describe the MSE of the reconstruction. Finally, requiring that the MSE vanishes provides the threshold for perfect recovery. The next section reports the outcomes of the analysis and the last part of the paper is devoted for describing how the calculations were carried out.

III. RESULTS

Let Q denote the standard Q-function and define two functions

$$s(x) = \frac{1}{x^2} [1 - 2Q(x)] - \sqrt{\frac{2}{\pi x^2}} e^{-\frac{x^2}{2}}, \quad (8)$$

$$r_{\lambda}(h) = \lambda \sqrt{\frac{h}{2\pi}} e^{-\frac{\lambda^2}{2h}} - (\lambda^2 + h) Q\left(\frac{\lambda}{\sqrt{h}}\right), \quad (9)$$

that take positive real arguments. Then, under the (technical) assumption that the extrema of the free energy falls within the replica symmetric ansatz (see, e.g., [12]–[14] for further discussion), the following results are derived in Section IV.

Proposition 1. Fix $\lambda, \alpha, \rho_x, \rho_w$ and let the variances σ_x^2 and σ_w^2 be finite and greater than zero. Then, the threshold for the perfect reconstruction, $\text{mse} \rightarrow 0$, is given by the solution of

$$A = \frac{\rho_x(\lambda^2 + \hat{\chi}) - 2(1 - \rho_x)r_{\lambda}(\hat{\chi})}{[2(1 - \rho_x)Q(\lambda/\sqrt{\hat{\chi}}) + \rho_x]^2}, \quad (10)$$

$$\hat{\chi} = \alpha(1 - \rho_w) \left\{ A \left[1 - 2Q\left(\frac{1}{\sqrt{A}}\right) \right] - \sqrt{\frac{2A}{\pi}} e^{-1/(2A)} + 2Q\left(\frac{1}{\sqrt{A}}\right) \right\} + \alpha\rho_w, \quad (11)$$

that satisfies the condition

$$\alpha(1 - \rho_w) \left[1 - 2Q\left(\frac{1}{\sqrt{A}}\right) \right] = (1 - \rho_x)2Q\left(\frac{\lambda}{\sqrt{\hat{\chi}}}\right) + \rho_x. \quad (12)$$

Proposition 2. Let the system be outside of the perfect reconstruction phase given by Proposition 1. The MSE of the

sparse signal estimate obtained with (2) and (3) is then

$$\text{mse} = \rho_x \sigma_x^2 - 4\sigma_x^2 \rho_x Q\left(\frac{\lambda}{\sqrt{\hat{\chi}} + \sigma_x^2 \hat{m}^2}\right) - 2\hat{m}^{-2}[(1 - \rho_x)r_\lambda(\hat{\chi}) + \rho_x r_\lambda(\hat{\chi} + \sigma_x^2 \hat{m}^2)], \quad (13)$$

where the required parameters can be obtained by solving

$$\chi = \frac{2}{\hat{m}} \left[(1 - \rho_x) Q\left(\frac{\lambda}{\sqrt{\hat{\chi}}}\right) + \rho_x Q\left(\frac{\lambda}{\sqrt{\hat{\chi}} + \sigma_x^2 \hat{m}^2}\right) \right], \quad (14)$$

$$\hat{m} = \frac{\alpha(1 - \rho_w)}{\chi} \left[1 - 2Q\left(\frac{\chi}{\sqrt{\text{mse}}}\right) \right] + \frac{\alpha\rho_w}{\chi} \left[1 - 2Q\left(\frac{\chi}{\sqrt{\text{mse} + \sigma_w^2}}\right) \right], \quad (15)$$

$$\hat{\chi} = \alpha(1 - \rho_w) \left[s\left(\frac{\chi}{\sqrt{\text{mse}}}\right) + 2Q\left(\frac{\chi}{\sqrt{\text{mse}}}\right) \right] + \alpha\rho_w \left[s\left(\frac{\chi}{\sqrt{\text{mse} + \sigma_w^2}}\right) + 2Q\left(\frac{\chi}{\sqrt{\text{mse} + \sigma_w^2}}\right) \right]. \quad (16)$$

The results imply that in the LSL the reconstruction has two phases; one where the MSE of the reconstruction vanishes and another where the estimate is noisy. The threshold is characterized by the parameters $\{\lambda, \alpha, \rho_x, \rho_w\}$ implying that the condition for perfect reconstruction is independent of the signal and noise variances σ_x^2 and σ_w^2 , respectively. Outside of the perfect reconstruction phase, however, the signal to noise ratio of the system has an impact on the MSE.

Mean square error predicted by Proposition 2 is shown in Fig. 1a. Numerical experiments obtained with `cvx` [6] are also given. Below the thresholds $\rho_x = 0.0770$ and $\rho_x = 0.1030$ for $\lambda = 1$ and $\lambda = \text{optimal}$, respectively, the MSE of the reconstruction vanishes. Figure 1b shows the effect of λ on the perfect recovery threshold given in Proposition 1. Here $\rho_w = \delta\rho_x$, where $\delta = 1/5, 1/10, 1/50$. For given ρ_x we find the critical α that admits perfect reconstruction, so that the MSE vanishes for the set of parameters that lie above the selected curve. The results demonstrate that the choice of parameter λ has a significant impact on the performance. Note that optimization of λ with simulations is extremely time consuming, while it is easy to do by using Proposition 1.

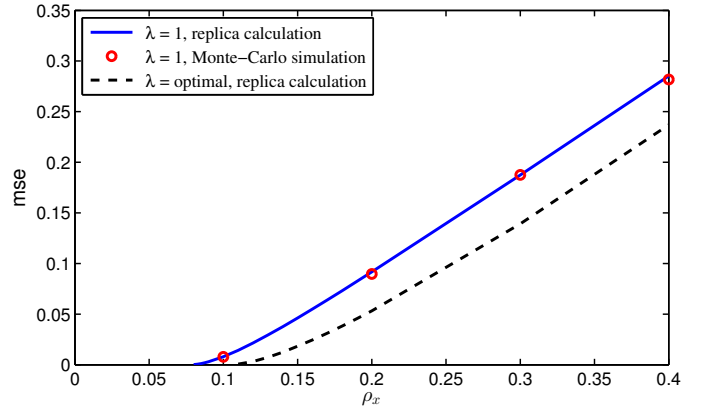
IV. REPLICA ANALYSIS

In this section a sketch of derivation is given for Propositions 1 and 2. Throughout the rest of the paper, the replica trick given in Section II-A is assumed to be valid. With this in mind, recall (7) and denote $v^a = \mathbf{A}(\mathbf{x}^0 - \mathbf{x}^a)$. The term inside \ln in (6) can then be written as

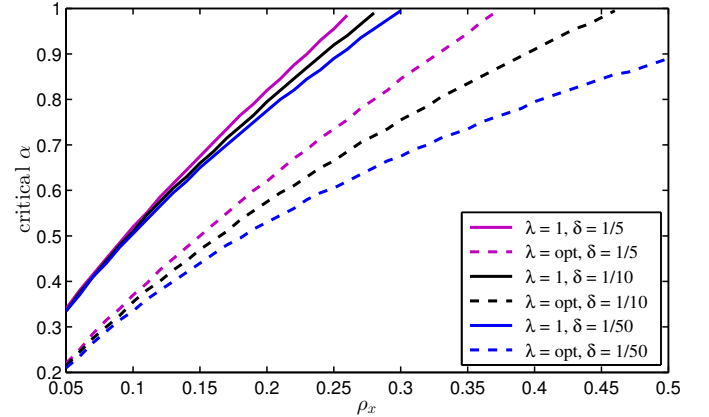
$$\mathbb{E}_{\mathbf{x}^0} \left\{ \int \prod_{a=1}^u \left[e^{-\beta\lambda\|\mathbf{x}^a\|_1} d\mathbf{x}^a \right] \mathbb{E}_{\mathbf{A}, \mathbf{w}} \prod_{a=1}^u e^{-\beta\|\mathbf{v}^a + \mathbf{w}\|_1} \right\}, \quad (17)$$

where \mathbf{x}^0 has IID elements drawn according to $p(x; \rho_x, \sigma_x^2)$. We first concentrate on evaluating the latter term $\mathcal{I}_{u,\beta}(\mathcal{X}) = \mathbb{E}_{\mathbf{A}, \mathbf{w}} \prod_{a=1}^u e^{-\beta\|\mathbf{v}^a + \mathbf{w}\|_1}$, for a fixed set $\mathcal{X} = \{\mathbf{x}^a\}_{a=0}^u$.

Since \mathbf{A} has IID elements with density $g_A(0, 1/N)$, conditioned on \mathcal{X} the vectors $\{\mathbf{v}^a\}$ tend to jointly Gaussian RVs



(a) MSE of reconstruction for $\alpha = 1/2$, $\rho_w = 0.1$ and $\sigma_x^2 = \sigma_w^2$. Lines for replica analysis based results and markers for simulations.



(b) Replica method based critical condition for perfect reconstruction. Solid lines for $\lambda = 1$ and dashed lines for optimal λ .

Fig. 1. Reconstruction performance vs. signal density ρ_x .

by the central limit theorem as $N \rightarrow \infty$. More precisely, if $\mathbf{v} \in \mathbb{R}^{uN}$ is formed by stacking $\{\mathbf{v}^a\}_{a=1}^u$ then \mathbf{v} is a zero-mean Gaussian RV with covariance matrix $\mathbf{R} = \mathbb{E}_A \mathbf{v} \mathbf{v}^T$. We write this as $\mathbf{v} \sim g_v(\mathbf{0}, \mathbf{R})$ and note that the (a, b) th $(a, b = 0, 1, \dots, u)$ block of \mathbf{R} is given by

$$\mathbf{R}_{a,b} = [Q_{00} - (Q_{a0} + Q_{0b}) + Q_{ab}] \mathbf{I}_M, \quad (18)$$

where $Q_{ab} = N^{-1}(\mathbf{x}^a \cdot \mathbf{x}^b)$. For later use, let $\mathbf{Q} \in \mathbb{R}^{(u+1) \times (u+1)}$ be composed of the elements $\{Q_{ab}\}$. Thus, for large N ,

$$\mathcal{I}_{u,\beta}(\mathcal{X}) = \mathbb{E}_w \int g_v(\mathbf{0}, \mathbf{R}) \exp \left[-\beta \sum_{a=1}^u \|\mathbf{v}^a + \mathbf{w}\|_1 \right] d\mathbf{v}, \quad (19)$$

where we have omitted terms that vanish as $N \rightarrow \infty$ [12], [13].

In the large system limit of $N \rightarrow \infty$, Laplace's (the saddle point) method with respect to \mathbf{R} yields the exact assessment of $N^{-1} \ln E\{[Z_\beta(\mathbf{y}, \mathbf{A}; \lambda)]^u\}$ for $\forall n \in \mathbb{N}$ and $\forall \beta > 0$. We here assume that the dominant saddle point in the assessment is invariant under any permutation of the replica indexes $a = 1, 2, \dots, u$, which is often termed the replica symmetric (RS) ansatz and is characterized as $Q_{a0} = Q_{0b} = m$, $Q_{aa} = Q$, $a = 1, \dots, u$ and $Q_{ab} = q$ for $a \neq b \in \{1, \dots, u\}$ in

the current case. This allows us to express \mathbf{v}^a in (19) as $\mathbf{v}^a = \mathbf{z}_a \sqrt{Q - q} + t \sqrt{p - 2m + q} \in \mathbb{R}^M$, which means that $\mathcal{I}_{u,\beta}(\mathcal{X})$ is proportional to

$$\left[\mathbb{E} \left\{ \left(\int e^{-\beta |z \sqrt{Q-q} + t \sqrt{p-2m+q} + w| - \frac{1}{2} z^2} dz \right)^u \right\} \right]^M, \quad (20)$$

where $\mathbb{E}\{\dots\} = \int (\dots) p(w; \rho_w, \sigma_w^2) dw Dt$ and $Dt = dt e^{-t^2/2} / \sqrt{2\pi}$. Since we are interested in the zero temperature solution $\beta \rightarrow \infty$, Laplace's method for the integral w.r.t. z implies

$$\mathcal{I}_{u,\beta}(\mathbf{Q}) \propto \left[\mathbb{E} e^{-u\beta \psi(t,w;\mathbf{Q})} \right]^{\alpha N}. \quad (21)$$

We write next the exponential term in (20) in a slightly different form by denoting $\chi = \beta(Q - q) \geq 0$ and $f(t, w; \mathbf{Q}) = t \sqrt{p - 2m + Q} + w$. We also use the fact that $p - 2m + q = p - 2m + Q - \chi/\beta \rightarrow p - 2m + Q$ for any finite χ . The Laplace's method requires then that

$$\psi(t, w; \mathbf{Q}) = \min_z \left\{ |z \sqrt{\chi} + f(t, w; \mathbf{Q})| + \frac{z^2}{2} \right\}. \quad (22)$$

Examining the critical points of ψ for a fixed set $\{t, w, \chi, \mathbf{Q}\}$ shows that the minimizing z gives

$$\psi(t, w; \mathbf{Q}) = \begin{cases} f(t, w; \mathbf{Q})^2 / (2\chi), & |f(t, w; \mathbf{Q})| < \chi; \\ |f(t, w; \mathbf{Q})| - \chi/2, & |f(t, w; \mathbf{Q})| > \chi. \end{cases} \quad (23)$$

The next task is to average $\mathcal{I}_{u,\beta}(w; \mathbf{Q})$ over the set \mathcal{X} . The expectation w.r.t. \mathbf{Q} can be carried out under the RS ansatz by defining first the probability weight

$$\mu(\mathbf{Q}) = \int d\mathbf{x}^0 p(\mathbf{x}^0) \prod_{a=1}^u \left(d\mathbf{x}^a e^{-\beta \lambda \|\mathbf{x}^a\|_1} \right) \times \prod_{0 \leq a \leq b \leq u} \delta(\mathbf{x}^a \cdot \mathbf{x}^b - N Q_{ab}), \quad (24)$$

and integrating then w.r.t. the measure $\mu(\mathbf{Q})$. Under the RS ansatz (24) has the same form as in [16], [17] so we skip the derivation here due to space constraints and arrive straight at the expression

$$\mu(\mathbf{Q}) \propto \int d\hat{\mathbf{Q}} \exp \left[\beta N \left(u \frac{\hat{Q}Q - \hat{\chi}\chi}{2} - um\hat{m} + \frac{u^2}{2} (\hat{\chi}\chi - \beta \hat{\chi}Q) + \frac{1}{\beta} \log \mathcal{M}_u(\hat{\mathbf{Q}}; \beta, \lambda) \right) \right], \quad (25)$$

where $\hat{\mathbf{Q}}$ is a short-hand for $\{\hat{\chi}, \hat{Q}, \hat{m}\}$. We also have the moment generating function of the elements of $\{\mathbf{x}^a\}_{a=0}^u$

$$\mathcal{M}_u(\hat{\mathbf{Q}}; \beta, \lambda) = (1 - \rho_x) \mathbb{E}_z e^{-\beta u \phi_\lambda(z \sqrt{\hat{\chi}}; \hat{Q})} + \rho_x \mathbb{E}_z e^{-\beta u \phi_\lambda(z \sqrt{\hat{\chi} + \sigma_x^2 \hat{m}^2}; \hat{Q})}, \quad (26)$$

where $\mathbb{E}_z(\dots)$ denotes $\int (\dots) Dz$ and ϕ satisfies

$$\phi_\lambda(h; \hat{Q}) = \begin{cases} -(|h| - \lambda)^2 / (2\hat{Q}), & \text{if } |h| > \lambda \\ 0, & \text{if } |h| \leq \lambda. \end{cases} \quad (27)$$

The final form of $\mu(\mathbf{Q})$ seems undoubtedly rather cryptic so we give here a few hints how it can be obtained (more details in [16], [17]). The first task in obtaining (25) is to write the Dirac's delta functions using (inverse) Fourier transform and integrating over \mathbf{x}^0 with the help of the Gaussian integral

$$\sqrt{\frac{1}{2\pi}} \int e^{-ax^2/2 + bx} dx = \frac{1}{\sqrt{a}} \exp\left(\frac{b^2}{2a}\right). \quad (28)$$

Then (28) is used right-to-left to decouple the replicated terms $\{\mathbf{x}^a\}$ and the average over them is obtained using the saddle point method as $\beta \rightarrow \infty$. These last two steps give rise to (27) and the integrals in (26). Rest of the terms in (25) arise essentially from the (inverse) Fourier transform of the Dirac's delta functions where the hatted variables represent scaled transform domain variables.

Combining (21) and (25) yields an expression for (17) as

$$\mathbb{E}\{[Z_\beta(\mathbf{y}, \mathbf{A}; \lambda)]^u\} \propto \int d\mathbf{Q} \mathcal{I}_{u,\beta}(\mathbf{Q}) \mu(\mathbf{Q}) \quad (29)$$

$$= \int d\mathbf{Q} d\hat{\mathbf{Q}} \exp \left\{ \beta N \left(\frac{\alpha}{\beta} \log \mathbb{E}_{t,w} e^{-u\beta \psi(t,w;\mathbf{Q})} - um\hat{m} + u \frac{\hat{Q}Q - \hat{\chi}\chi}{2} + \frac{u^2}{2} (\hat{\chi}\chi - \beta \hat{\chi}Q) + \frac{1}{\beta} \log \mathcal{M}_u(\hat{\mathbf{Q}}; \beta, \lambda) \right) \right\}.$$

For the integration w.r.t. \mathbf{Q} and $\hat{\mathbf{Q}}$ we use again the saddle point method as $N \rightarrow \infty$. Note that we have then by the law of large numbers $p \rightarrow \sigma_x^2 \rho_x$ as well (see (20)). Thus, the replica symmetric expression for (6) reads

$$f_{rs}(\lambda) = \text{extr} \left\{ \frac{\hat{Q}Q}{2} - \frac{\hat{\chi}\chi}{2} - m\hat{m} + \lim_{u \rightarrow 0^+} \frac{\partial}{\partial u} \left[\frac{\alpha}{\beta} \log \mathbb{E} e^{-u\beta \psi(t,w;\mathbf{Q})} + \frac{1}{\beta} \log \mathcal{M}_u(\hat{\mathbf{Q}}; \beta, \lambda) \right] \right\}, \quad (30)$$

where we used the fact that the order of extremization $\text{extr}\{\dots\}$ w.r.t. $\{\chi, m, Q, \hat{\chi}, \hat{m}, \hat{Q}\}$ and the partial derivative w.r.t. u can be exchanged [12]. Solving the remaining derivatives finally gives the form

$$f_{rs}(\lambda) = \text{extr} \left\{ \frac{\hat{Q}Q}{2} - \frac{\hat{\chi}\chi}{2} - m\hat{m} - \alpha \mathbb{E}_{t,w} \psi(t, w; \chi, m, Q) - (1 - \rho_x) \mathbb{E}_z \phi_\lambda(z \sqrt{\hat{\chi}}; \hat{Q}) - \rho_x \mathbb{E}_z \phi_\lambda(z \sqrt{\hat{\chi} + \sigma_x^2 \hat{m}^2}; \hat{Q}) \right\}, \quad (31)$$

in the limit $u \rightarrow 0^+$.

We have now managed to write the normalized free energy under RS ansatz as the solution of an extremization problem that has a couple of expectations inside. Let us first consider the derivatives w.r.t. the "hat-variables" $\{\hat{\chi}, \hat{m}, \hat{Q}\}$. Since ψ does not depend on them, the only non-trivial part is to solve the expectations and partial derivatives on the second line in (31).

Lemma 1. Let h be a real positive (function) independent of $z \in \mathbb{R}$. Then, for positive real parameters \hat{Q} and λ we have

$$\mathbb{E}_z \phi_\lambda(z\sqrt{h}; \hat{Q}) = \hat{Q}^{-1} r_\lambda(h), \quad (32)$$

$$\frac{\partial}{\partial x} r_\lambda(h) = -\left(\frac{\partial h}{\partial x}\right) Q\left(\frac{\lambda}{\sqrt{h}}\right), \quad (33)$$

where $r_\lambda(h)$ is given in (9).

Using the above results, the normalized free energy reads

$$f_{rs}(\lambda) = \text{extr} \left\{ \frac{\hat{Q}Q}{2} - \frac{\hat{\chi}\chi}{2} - m\hat{m} - \alpha \mathbb{E}_{t,w} \psi(t, w; \chi, m, Q) \right. \\ \left. + \hat{Q}^{-1} [(1 - \rho_x) r_\lambda(\hat{\chi}) + \rho_x r_\lambda(\hat{\chi} + \sigma_x^2 \hat{m}^2)] \right\}, \quad (34)$$

where χ is given in (14) and

$$m = 2\sigma_x^2 \rho_x \left(\frac{\hat{m}}{\hat{Q}} \right) Q\left(\frac{\lambda}{\sqrt{\hat{\chi} + \sigma_x^2 \hat{m}^2}} \right), \quad (35)$$

$$Q = -2\hat{Q}^{-2} [(1 - \rho_x) r_\lambda(\hat{\chi}) + \rho_x r_\lambda(\hat{\chi} + \sigma_x^2 \hat{m}^2)]. \quad (36)$$

Lemma 2. Let $f(t\sqrt{a}, x_1, \dots, x_k)$ be a real-valued function, where $a \geq 0$ and $\{t, x_1, \dots, x_k\}$ are independent random variables that do not depend on a . Then,

$$\frac{\partial}{\partial a} \int f(t\sqrt{a}, x_1, \dots, x_k) Dt = \frac{1}{2} \int f''(t\sqrt{a}, x_1, \dots, x_k) Dt, \quad (37)$$

where $f''(\dots)$ is the 2nd order partial derivative w.r.t. first argument. Also, denoting the indicator function $\mathbb{1}\{\dots\}$,

$$\int \mathbb{1}\{|t| > a\} Dt = 2Q(a), \quad (38)$$

$$\int \mathbb{1}\{|t| < a\} Dt = 1 - 2Q(a), \quad (39)$$

$$\int t^2 \mathbb{1}\{|t| < a\} Dt = 1 - 2Q(a) - \frac{2a}{\sqrt{2\pi}} e^{-a^2/2}, \quad (40)$$

where the integrals are over the set of real numbers.

Using (37) for the partial derivatives w.r.t. m and Q , and then (38) and (40) for the remaining integrals shows that $\hat{Q} = \hat{m}$ as given in (15). Furthermore, $\text{mse} = \sigma_x^2 \rho_x - 2m + Q$ reduces to (13) and gives the MSE of the reconstruction [16], [17]. Similarly, from the derivative of χ and (38) – (40) one gets (16). Thus, we have obtained a full description of the free energy under the RS ansatz in terms of six parameters. More importantly, we obtained as a by product the MSE behavior of the convex optimization problem based on (2) and (3), finishing the proof of Proposition 2.

To derive Proposition 1, we first require that $\text{mse} \rightarrow 0$, which implies $\rho_x \sigma_x^2 = m = Q$ and $\hat{m} = \hat{Q} \rightarrow \infty \implies \chi \rightarrow 0$. For a non-trivial solution we should have $\hat{\chi} \in O(1)$ and $0 < \lambda < \infty$. Using the Taylor expansion of the Q-function and exponential function near zero, the proposition follows after some algebra (details omitted due to space constraints).

REFERENCES

- [1] A. Miller, *Subset Selection in Regression (second edition)*. Chapman and Hall/CRC, 2002.
- [2] L. Daudet, “Sparse and structured decompositions of signals with the molecular matching pursuit,” *IEEE Trans. Audio, Speech, and Lang. Proc.*, vol. 14, no. 5, pp. 1808–1816, 2006.
- [3] J. Wright et al., “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.
- [4] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [5] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Proc. Mag.*, vol. 25, pp. 21–30, 2008.
- [6] CVX Research, Inc., “CVX: Matlab software for disciplined convex programming,” <http://cvxr.com/cvx>.
- [7] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [8] R. E. Carrillo, K. E. Barner, and T. C. Aysal, “Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise,” *IEEE J. Select. Topics in Signal Processing*, vol. 4, no. 2, pp. 392–408, 2010.
- [9] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, “Exact signal recovery from sparsely corrupted measurements through the pursuit of justice,” in *Proc. Annual Asilomar Conf. Signals, Syst., Comp.*, Nov., pp. 1556–1560.
- [10] A. Cherian, S. Sra, and N. Papanikolopoulos, “Denoising sparse noise via online dictionary learning,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 2060–2063.
- [11] V. Dotsenko, *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge University Press, 2001.
- [12] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Inform. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [13] D. Guo and S. Verdú, “Randomly spread CDMA: Asymptotics via statistical physics,” *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.
- [14] B. M. Zaidel, R. R. Müller, A. L. Moustakas, and R. De Miguel, “Vector precoding for Gaussian MIMO broadcast channels: Impact of replica symmetry breaking,” *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1413–1440.
- [15] K. Takeuchi, M. Vehkaperä, T. Tanaka, and R. Müller, “Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems,” *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1385–1412, 2012.
- [16] Y. Kabashima, T. Wadayama, and T. Tanaka, “A typical reconstruction limit for compressed sensing based on l_p -norm minimization,” *J. Stat. Mech.*, vol. 2009, no. 9, p. L09003, 2009.
- [17] Y. Kabashima, M. Vehkaperä, and S. Chatterjee, “Typical l_1 -recovery limit of sparse vectors represented by concatenations of random orthogonal matrices,” *J. Stat. Mech.*, no. 12, p. P12003, 2012.
- [18] S. Rangan, A. K. Fletcher, and V. K. Goyal, “Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 58, no. 3, pp. 1902–1923, 2012.
- [19] A. Tulino, G. Caire, S. Verdú, and S. Shamai, “Support recovery with sparsely sampled free random matrices,” *IEEE Trans. Inform. Theory*, to appear.